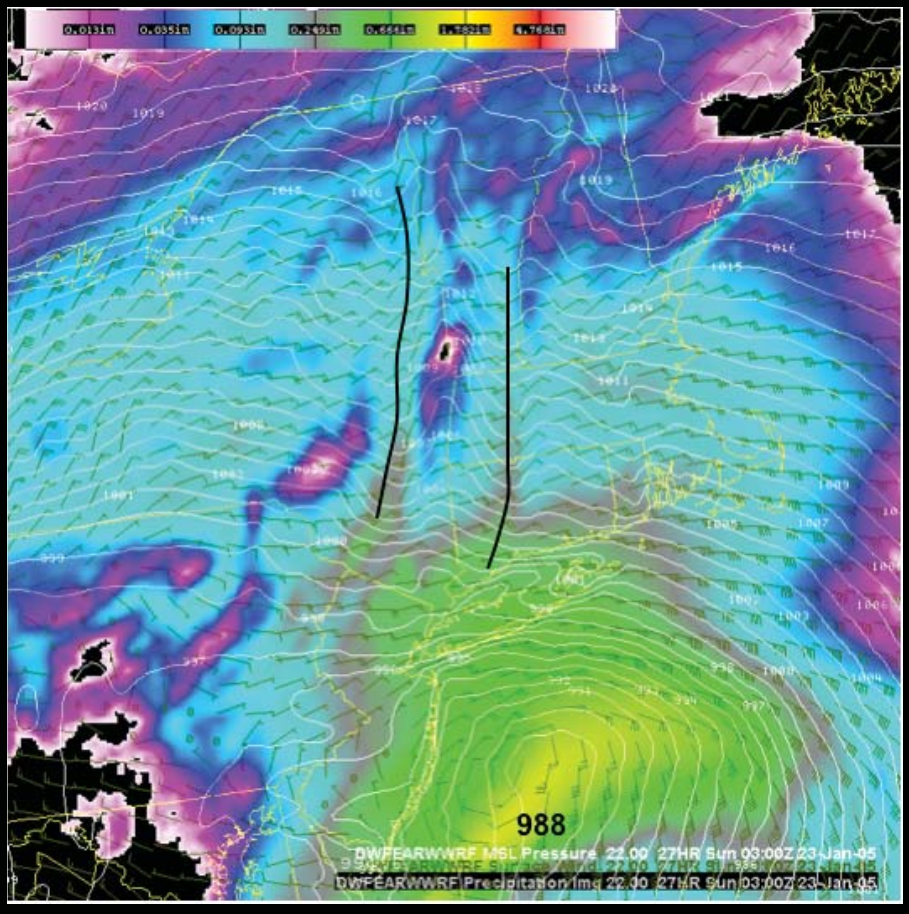
THE DEVELOPMENTAL TESTBED CENTER AND ITS WINTER FORECASTING EXPERIMENT

BY LÍGIA BERNARDET,* LOUISA NANCE, MERAL DEMIRTAS, STEVE KOCH, EDWARD SZOKE,+ TRESSA FOWLER, ANDREW LOUGHE,# JENNIFER LUPPENS MAHONEY, HUI-YA CHUANG, MATTHEW PYLE, AND ROBERT GALL

In a real-time experiment, operational and research meteorologists collaborated to evaluate usage of two configurations of a state-of-the-art numerical weather prediction model.



For the severe New England blizzard of 23 January 2005, 27-h forecasts of two versions of the WRF model (see text) show pronounced north-south bands north of the storm center in MSLP (contours) and three-hour accumulated precipitation (see color key) fields. These mesoscale bands (highlighted by north-south lines) were fixed to the terrain (the Hudson and Connecticut River Valleys separated by the Adirondack and Berkshire Mountains) for 9h during the peak of the storm. The observations agreed with this prediction, demonstrating the value of running high-resolution models over a large domain. Maximum precipitation forecast of ~2.0 in / 3h is shown east of the New Jersey coast.

The USWRP (see appendix B for all acronym expansions) includes plans for the rapid and direct transfer of new research results into operational NWP by the NWS and other agencies. The WRF model program is an important component of the USWRP technology transfer plan. WRF incorporates a community model infrastructure to facilitate scientific collaboration. The WRF effort embodies the concept of the operational and research communities working jointly toward the development of next-generation NWP capabilities by giving the research community better access to operational codes for testing and investigating forecast failures. This way, as new technologies are developed, the most promising results can be rapidly and efficiently transferred to operations.

To rapidly transfer new technology into operations, the WRF program must ensure accurate ▶

predictions and code integrity through controlled testing and evaluation over many cases. While all NWP centers have an operational testing center, where mature codes are tested prior to operational implementation, there is a lack of resources for testing new codes before they are ready for operational use. The academic community cannot routinely do such testing, because it requires supporting infrastructure and heavy computational resources. To fill this void, the WRF DTC was created in 2003, which is a facility where the NWP research and operational communities interact to test and evaluate new developments in the WRF model, for both research applications and operational implementation. The WRF DTC has staff and substantial computer resources distributed over a variety of nodes. Currently, there are two nodes in Boulder, Colorado—one at NCAR and the second at NOAA/ESRL. A third node is planned in Monterey, California, at the Naval Research Laboratory. Staff from the various nodes works closely with each other, as well as with the research and operational communities.

A WRF DTC's goal is to be the focal point for a diverse team of research and operational experts working on innovative ideas for NWP and implementing the best new options into the forecast process. A strong link between the research and operational communities increases the number of creative minds working on operational NWP improvements, leading to more rapid improvements in forecast accuracy. Membership of the WRF DTC Advisory Board is divided equally between representatives of the research and operational entities. The board assists the WRF DTC management in identifying emerging technologies for testing, reviewing proposals for the visitor program, and giving feedback to assist the WRF DTC management with high-level priorities.

The computational environment of the WRF DTC is functionally similar to that used in operations. Considerable effort was spent migrating meteorological software packages from NCEP to the WRF DTC, so that testing and evaluation of the WRF modeling system would not interrupt day-to-day forecasting operations.

The DWFE, a model evaluation exercise, exemplifies the interaction between the operational and research communities that is fostered by the WRF DTC. The objectives of DWFE were to generate WRF forecasts without cumulus parameterization on a grid with 5-km horizontal spacing over the CONUS, provide the model products on operational display and analysis systems to NWS forecasters, and determine the value of such high-resolution WRF models for winter weather forecasting.

This paper summarizes the WRF DTC activities, discusses DWFE with details about the experiment configuration, shows examples of objective verification and the potential use of DWFE products in the forecasting process, and evaluates the experiment.

THE DEVELOPMENTAL TESTBED CENTER.

Since 2003, the WRF DTC has performed two retrospective tests (the WRF Test Plan and WRF Rapid Refresh Core Test) and two real-time forecasting experiments (DWFE and NMM5-CONUS). All experiments made use of the WRF DTC end-to-end system, which is comprised of input data collection and preprocessing, WRF forecasting, postprocessing with the NCEP WPP, dissemination of forecasts, forecast verification, and archiving. In addition to the participation in forecasting experiments, the WRF DTC interacts with the research community, primarily through the WRF DTC Visitor Program and support of the WRF code.

Forecasting experiments. The first forecast evaluation project conducted by the WRF DTC was the WRF Test Plan (Seaman et al. 2004; Davis et al. 2004; Bernardet et al. 2004). This collaboration between scientists from NCAR, NOAA/ESRL, NCEP, and the Air Force Weather Agency examined a range of WRF model configurations for possible implementation at NCEP in a high-resolution WRF ensemble prediction system. This test domain corresponded to the NCEP

AFFILIATIONS: Bernardet, Koch, Szoke, Loughe, and Mahoney—NOAA/Earth System Research Laboratory, Boulder, Colorado; Nance, Demirtas, Fowler, and Gall—National Center for Atmospheric Research, Boulder, Colorado; Chuang and Pyle—Environmental Modeling Center, National Centers for Environmental Prediction, Silver Spring, Maryland
*__ADDITIONAL AFFILIATION:__ Systems Research Group, Colorado Springs, Colorado
+__ADDITIONAL AFFILIATION:__ Cooperative Institute for Research in the Atmosphere, Fort Collins, Colorado
#__ADDITIONAL AFFILIATION:__ Cooperative Institute for Research in Environmental Sciences, Boulder, Colorado
CORRESPONDING AUTHOR: Dr. Lígia Bernardet, NOAA/Earth System Research Laboratory, 325 Broadway GSD7, Boulder, CO 80305
E-mail: ligia.bernardet@noaa.gov

The abstract for this article can be found in this issue, following the table of contents.
DOI:10.1175/BAMS-89-5-611

In final form 21 November 2007

HRWs, using grid spacing ranging from 8 to 10 km. Eight different configurations of WRF were tested, with variations in initial conditions (unperturbed conditions directly interpolated from a parent model, or conditions perturbed by bred anomalies), physics suites (NCAR and NCEP packages), and dynamic cores. Two cores within the WRF software framework were tested: the NMM developed by NCEP (Janjic 2003) and the ARW developed by NCAR (Skamarock et al. 2005). Each configuration was applied to a full month from each season to test the model in a variety of weather regimes. Although the planned 8-km grid spacing, high-resolution ensemble was never implemented at NCEP because of a lack of computer resources, the results of this test indicated that both WRF cores were ready to be transitioned to operations, leading in September 2004 to the implementation of one ARW and one NMM configuration of the WRF in all HRWs. Results from the test plan were used in extended analyses by WRF DTC visitors from the research community (Gallus and Bresch 2006) and by developers at NCEP. The latter effort resulted in the development of an experimental version of the NCEP convective parameterization (Janjic 2004) that is capable of generating fields with greater magnitudes and finer-scale structure than the original parameterization.

The next forecasting experiment conducted by WRF DTC was DWFE, a real-time experiment that took place from 15 January through 31 March 2005. DWFE's goals were to

- provide experimental model guidance for winter weather forecasting over a large domain using two variants of the WRF model run at high resolution with only explicit convection (no convective parameterization scheme);
- expose forecasters to the nature and behavior of the WRF modeling system prior to the 2006 scheduled implementation of the WRF model in the NAM domain at NCEP;
- use both subjective and objective verification methods to determine whether the encouraging results seen in earlier 4-km grid spacing WRF configurations that provided warm season forecast values (Done et al. 2004) extend to the winter for lead times out to 48 h; and
- determine the extent to which various mesoscale phenomena important to winter weather forecasting could be skillfully predicted.

One outcome of DWFE was that NWS forecasters participating in the project were very interested in the new products, and requested an extension of the forecasts into spring 2005. This request was accommodated through the WRF DTC NMM5-CONUS project, which extended the dissemination of the NMM forecasts until the end of July 2005. The results of NMM5-CONUS were also intended to support the NOAA Coastal Storms Initiative to address the impact of model domain size on local forecasting. The NMM5-CONUS forecasts were compared to those made with an identical version of the NMM for a much smaller domain centered over Florida. This study examined the influence of boundary conditions on short-range numerical weather predictions (out to 48 h), in order to determine whether the same degree of accuracy could be obtained with a much smaller domain. Interestingly, results showed that the use of the large domain offered no significant advantage over the small domain, pointing to the good quality of the boundary conditions provided by the Eta Model to the WRF model configured over the small domain (Bernardet et al. 2005).

The most recent testing activity of the WRF DTC, completed in July 2006, was the WRF Rapid Refresh Core Test (Nance 2006), a collaboration with the NOAA/ESRL Global Systems Division and the NCAR Research Applications Laboratory. This test was designed to fulfill the need for a controlled comparison of the two dynamic solvers available in WRF and to assist in the decision of which numerical core to use in the WRF Rapid Refresh, which is the model that will replace the Rapid Update Cycle model at NCEP in 2009. The controlled comparison of the two WRF dynamic cores was made possible by extensive work in creating physics interoperability between the cores. The ARW and NMM cores, configured on virtually identical 13-km grid spacing domains covering the CONUS, were run from the same initial and boundary data, and they used the same set of physics. Retrospective forecasts were initialized twice a day and run out to 24 h for a total of 4 months spanning four seasons. This test led to the recommendation by ESRL (Brown et al. 2007), and subsequent acceptance by NCEP, of the ARW dynamic core for initial operational implementation in the WRF Rapid Refresh.

Interactions with the academic community. The WRF DTC Visitor Program was initiated in the summer of 2004 with 3 visitors, and expanded in 2005 to 11 participants from universities, NCAR, NCEP, and the Joint Center for Satellite Data Assimilation. These visitors pursued a number of topics, such as data assimilation techniques, physics package testing,

examination of the dynamic cores of WRF, development of new verification techniques, and analysis of the forecasting experiments and tests conducted by the WRF DTC. In 2006, the WRF DTC Visitor Program held a well-attended DTC Visitor Reunion in Boulder (DTC 2006), and in 2007 another 10 visitors were funded. The WRF DTC intends to continue to maintain the Visitor Program.

In another effort to bridge the gap between research and operations, the WRF DTC is now supporting the distribution of the NMM core and of the WPP to the general community. The center offered five NMM tutorials between 2005 and 2007 to foster community access to the codes run operationally at NCEP so that forecasts can be evaluated and systematic problems can be identified and addressed. This effort stemmed from close interactions with NMM developers to prepare a user's guide and documentation, and involved preliminary testing of NMM and its pre- and postprocessing systems on various computing platforms. This effort parallels NCAR's support of the ARW model. Ongoing support of both cores is available online (see http://wrf-model.org).

WRF DTC WINTER FORECASTING EXPERIMENT. *End-to-end forecast system.* The high-resolution WRF forecasts produced during DWFE employed the two dynamic cores available within the WRF software framework: ARW and NMM. Once-daily (0000 UTC cycle), 48-h forecasts were generated for each core using a physics suite recommended by the developers of the respective dynamic cores. Both setups are shown in Table 1. Note that the dynamic cores use different grid staggering and map projections, so while the domains cover nearly the same area, it is impossible to make the domains identical. Information on the Eta Model (Black 1994), NCEP's primary short-term prediction system running in the NAM window at the time, is included for comparison. Note that the physics packages used by the Eta Model and NMM are a close match, except that the Eta Model used a convective parameterization scheme while the NMM did not. On the other hand, the two WRF configurations used a grid spacing of 5 km, while the Eta Model used a grid spacing of 12 km.

The initial and boundary conditions for both cores were based on the Eta Model's 40-km grid spacing output, known as Eta 212, processed using the WRF standard initialization. In addition to the Eta 212 data, the initial land surface fields for the ARW were obtained from the HRLDAS (Chen et al. 2004). HRLDAS utilizes observations and land characteristics to drive the Noah LSM in an uncoupled mode to capture finescale heterogeneity in land state. HRLDAS is processed on the same grid used to run the WRF model, allowing assimilated land state variables to be ingested directly into the coupled WRF–Noah LSM forecast system without interpolation.

Table 1. Configurations of the ARW, NMM, and Eta Model during DWFE.

	ARW	NMM	Eta
Land surface model	Noah 5 layer*	Noah 5 layer*	Noah 5 layer*
Boundary layer	Yonsei University	Mellor–Yamada–Janjic 2.5	Mellor–Yamada–Janjic 2.5
Microphysics	WRF Simple Microphysics 5	Ferrier	Ferrier
Cumulus parameterization	None	None	Betts–Miller–Janjic
Shortwave radiation	Dudhia	Lacis–Hansen	Lacis–Hansen
Longwave radiation	RRTM	Fels-Schwartzkopf GFDL	Fels–Schwartzkopf GFDL
Projection	Lambert conformal	Rotated lat–lon	Rotated lat–lon
Grid staggering	C	E	E
Vertical coordinate	Terrain-following sigma	Hybrid: terrain-following sigma and isobaric	Step mountain
Horizontal grid spacing	5 km	5 km	12 km
Number of verticals levels	37	37	60
Time step	30 s	10 s	30 s
Initial conditions	Eta 212 + HRLDAS	Eta 212	Eta Data Assimilation System
Boundary conditions	Eta 212	Eta 212	GFS

*The version of the Noah LSM is not identical in the ARW, NMM, and Eta models.

For this experiment, the NMM forecasts were generated using NOAA/ESRL's supercomputer, which is a Pentium IV Linux cluster, while the ARW core was run on NCAR's IBM SP cluster system. The computational requirements of DWFE were very large and placed an unprecedented demand on the computer systems. A sophisticated computer infrastructure had to be used to control the end-to-end forecast system, leading to the development of a workflow manager to automate DWFE model runs at ESRL. Details of postprocessing and forecast verification methodologies used in DWFE are described in appendix A.

In order to reach operational forecasters and the community at large, the WRF DTC invested resources into the development of the following four types of real-time displays to disseminate DWFE output:

- The NMM forecasts for a select subset of two-dimensional fields (primarily surface and precipitation fields) were made available through the AWIPS in the Central, Eastern, and Southern Region NWS WFOs. A limited number of WFOs used AWIPS, however, because of severe bandwidth constraints in the local forecast offices.
- Images for a select number of fields and levels were generated with NCAR Command Language and made available through the WRF DTC Web site (online at www.DTCenter.org). These images, as well as corresponding radar composites and Eta Model forecasts, are available via the DWFE catalog hosted by the NCAR EOL (online at http://catalog.eol.ucar.edu/dwfe).
- User-specified images were made available through FX-Net, an AWIPS-like interface that allows the user to interrogate the full three-dimensional model grids. FX-Net uses a client–server protocol to generate images on demand and to efficiently transfer the images to the users' display using a wavelet compression technique (Wang et al. 2002).
- Three-dimensional, postprocessed GRIB files were available to registered users, such as the NCEP Hydrometeorological Prediction Center.

A DWFE archive is being maintained by the WRF DTC on the NCAR Mass Store System. The archives are comprised of the WRF forecasts on the native and postprocessed grids, verification results, and observations used for verification. The archives are open to the community and access can be solicited through the WRF DTC Web site. Additionally, selected images can be viewed at the EOL catalog (online at http://catalog.eol.ucar.edu/dwfe/).

Examples of objective verification results. PRECIPITATION. The frequency bias scores computed using the NCEP QPFV system (a grid-to-grid approach in which forecast and observations are both interpolated to the 12-km G218) are shown in Fig. 1a. The WRF models overpredicted the area of precipitation at all thresholds, while the Eta Model overpredicted for thresholds smaller than 0.75 in. and underpredicted at higher thresholds.

The overprediction displayed by the WRF models is sensitive to the threshold chosen. In general, higher thresholds display higher overprediction. Uncertainty in the bias measures, represented by the 95% CI in the figures, is larger for higher thresholds because of the limited number of events at those thresholds. Although the ARW has smaller overprediction than the NMM, the differences are not significant, even for the highest thresholds. Note that the NMM bias scores are more similar to the ARW scores than to the Eta Model, even though the NMM physics package is more similar to that of the Eta

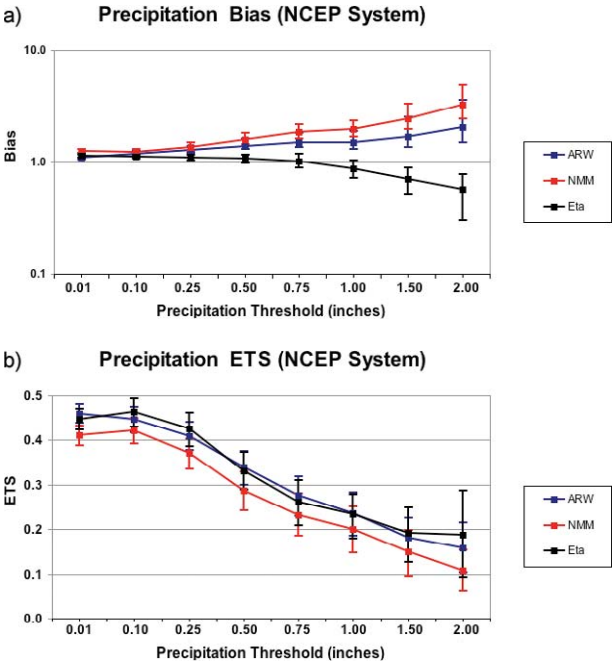


FIG. 1. Precipitation (a) bias and (b) ETS for the 24-h accumulation, averaged over the CONUS domain and over the entire DWFE period, computed using the NCEP system for several thresholds. Ninety-five-percent confidence intervals are included. ARW (blue), NMM (red), and Eta Model (black) are shown. Note that the ordinate axis of the bias plots follows a logarithmic scale. This representation was chosen in order to equally weigh over- and underprediction [i.e., the skill of a forecast whose areal coverage is half that observed (bias of 0.5) is equivalent to that of a forecast whose areal coverage is twice that observed (bias of 2.0)].

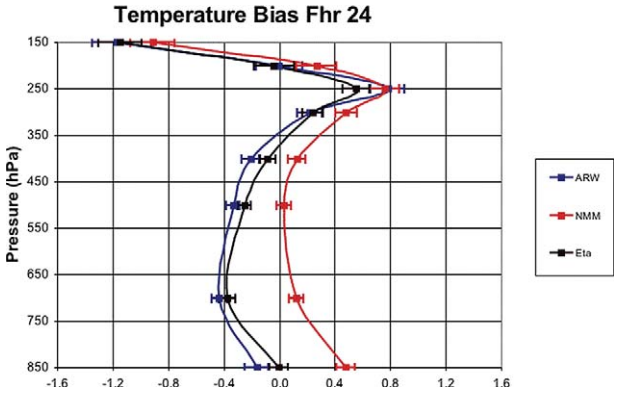


Fig. 2. Vertical profile for the 24-h temperature bias (°C) averaged over the CONUS domain and over the entire DWFE period. Ninety-five-percent confidence intervals are included. The ARW (blue), NMM (red), and Eta Model (black) are shown.

Model (Table 1). This behavior is possibly due to the similarity in grid spacing between ARW and NMM during DWFE and/or to the absence of cumulus parameterization in both WRF configurations. Models configured without cumulus parameterization require that saturation be reached on a grid cell before moist processes can occur. This requirement can delay the onset of convection, potentially leading to the development of large instabilities and excessive rainfall. The introduction of a positive-definite moisture advection scheme in a later version of the ARW core, WRF version 2.2 released in 2006, has been shown to alleviate the overprediction of rainfall in the ARW by avoiding the addition of moisture associated with the correction of negative values created by the Runge–Kutta advection scheme used in the ARW during DWFE. The lower resolution of the Eta Model, combined with the use of the BMJ cumulus parameterization, might be responsible for the low bias behavior of the Eta Model. Other investigations (Jankov et al. 2005) have shown that the BMJ parameterization tends to oversmooth the precipitation field and inflate the area covered by light precipitation at the expense of the representation of the high-precipitation amounts.

Differences in bias scores between the two WRF configurations may be attributed to a variety of sources, including differences in initial conditions (ARW uses HRLDAS, and NMM does not), the dynamic core, and physical parameterizations. It is not possible to determine the origin of the differences until further studies are completed. Jankov et al. (2005) discuss the sensitivity of warm-season precipitation forecasts to choices of physics packages, initial conditions, and the dynamic core. They show that bias scores are highly sensitive to those choices, and that the WRF model can generate either over- or underprediction of precipitation depending upon the configuration adopted.

Figure 1b shows the ETS computed using the NCEP QPFV system. In general, ETS decreases with threshold, indicating that the models have more difficulty in correctly forecasting the location of the higher-precipitation events. The results obtained for the NMM, ARW, and Eta Model are essentially similar, because the CIs overlap for the models. Hence, conclusions cannot be drawn regarding the superiority of any given model. The grid-to-station method, used in the RTVS, produced similar bias and ETS results to the method used by NCEP (shown in Demirtas et al. 2005b), with differences smaller than the uncertainty in the scores.

Temperature. Overall, for model forecasts of 24 h, the Eta Model and ARW had similar CONUS-averaged profiles of temperature bias, whereas the NMM was generally warmer than the other two models by approximately 0.4°C (Fig. 2). The NMM bias was positive at all levels below 150 hPa, reaching a maximum bias of nearly 0.8°C at 250 hPa. The Eta Model and ARW had negative biases at all levels (maximum of -1.2°C at 150 hPa) except 300 and 250 hPa, where positive biases did not exceed 0.8°C. The CIs indicate that the warm behavior of the NMM is statistically distinct from the other two models at all levels up to 300 hPa. In addition, the CIs indicate that the ARW model differs from the Eta Model at 850 and 400 hPa.

This pattern of temperature bias changed little from the initial time (Fig. 3) to later times in the forecast, indicating that biases present at the initial time persisted throughout the forecast. The reason for the differences between ARW and NMM temperature biases at the initial time can be attributed to differ-

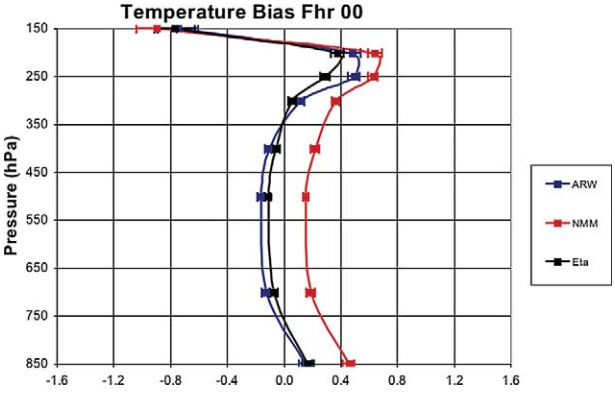


Fig. 3. Same as Fig. 2, except for the 0-h forecast.

ences in the initialization procedure for each core. Both models started from the same analysis; however, the ARW obtained temperature directly from the Eta 212 temperature field, whereas the NMM retrieved temperature from the Eta 212 geopotential using the hydrostatic approximation. Since DWFE, NMM developers have made a change in the initialization procedure and the current release of WRF is initialized from NAM temperatures for both cores.

A case study: The 01 March 2005 nor'easter. The nor'easter that affected the East Coast on 28 February and 1 March 2005 illustrates some of the value added by the DWFE high-resolution products. This storm deepened off the South Carolina coast and took the classic path up the East Coast to a position south of New England 24 h later (Fig. 4). It left behind a swath of snow accumulations ranging from 10 to 30 cm (4–12 in.), from Washington, D.C., and its western suburbs, northward across the Northeast (not shown).

Figure 5 illustrates the complex mesoscale structure of the snowstorm, which comprised four snowbands. The most coherent, long-lasting, and strongest of the snowbands was A, which propagated northward and then northwestward as the storm system developed. Bands B and C propagated tranversely to the dominant southwesterly flow aloft, and converged into a single mass sometime after 0900 UTC. Band D, which formed later in the life cycle of the storm, was a very organized transverse band. Thus, three of the four bands had the character expected of gravity waves, similar in appearance to precipitation bands seen in other strong winter storms affecting this region (e.g., Bosart and Sanders 1986; Zhang et al. 2001).

One of the new fields that was available during DWFE, and that proved to be very popular with forecasters, was simulated composite radar reflectivity (maximum reflectivity in a grid column). The 3-, 6-, 9-, and 12-h ARW and NMM forecasts of reflectivity (Fig. 6) can be compared against observed reflectivity (Fig. 5). Even though the NMM model displays a tendency to exhibit greater coverage of low reflectivity, it is interesting to note that the forecasts by the two WRF configurations are very similar, despite their numeric schemes and physical parameterizations being quite different. In particular, the placement, orientation, and number of precipitation bands forecast by the two models are remarkably similar to one another. Comparison between the forecasted and observed bands shows fairly good overall correspondence, particularly for bands A and D. However,

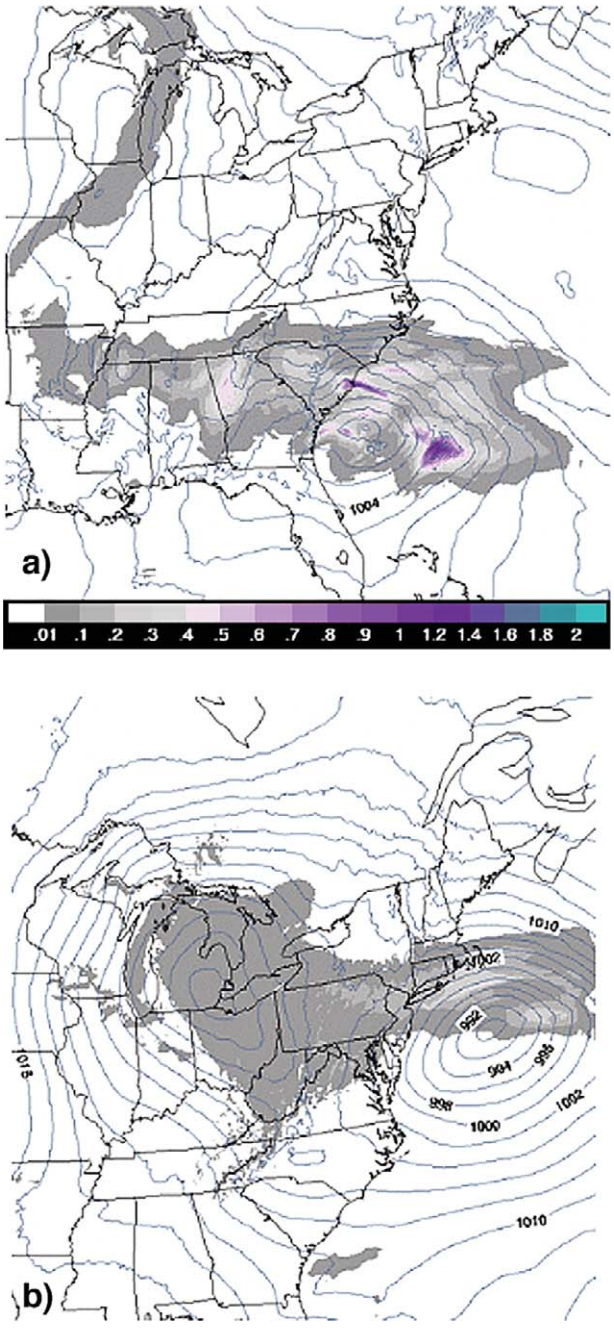


Fig. 4. Three-hour forecasts of mean sea level pressure (hPa) and 3-h accumulated precipitation (in.) from (a) the ARW valid at 0300 UTC 28 Feb and (b) the NMM valid at 0300 UTC 1 Mar 2005.

the details pertaining to the other two bands are only moderately well forecast.

A chief advantage of the simulated reflectivity product is that it allows one to more easily see detailed mesoscale structures forecast by high-resolution models, such as lake-effect snowbands, the structure of deep convection, and precipitation bands. The ability to discern these mesoscale bands

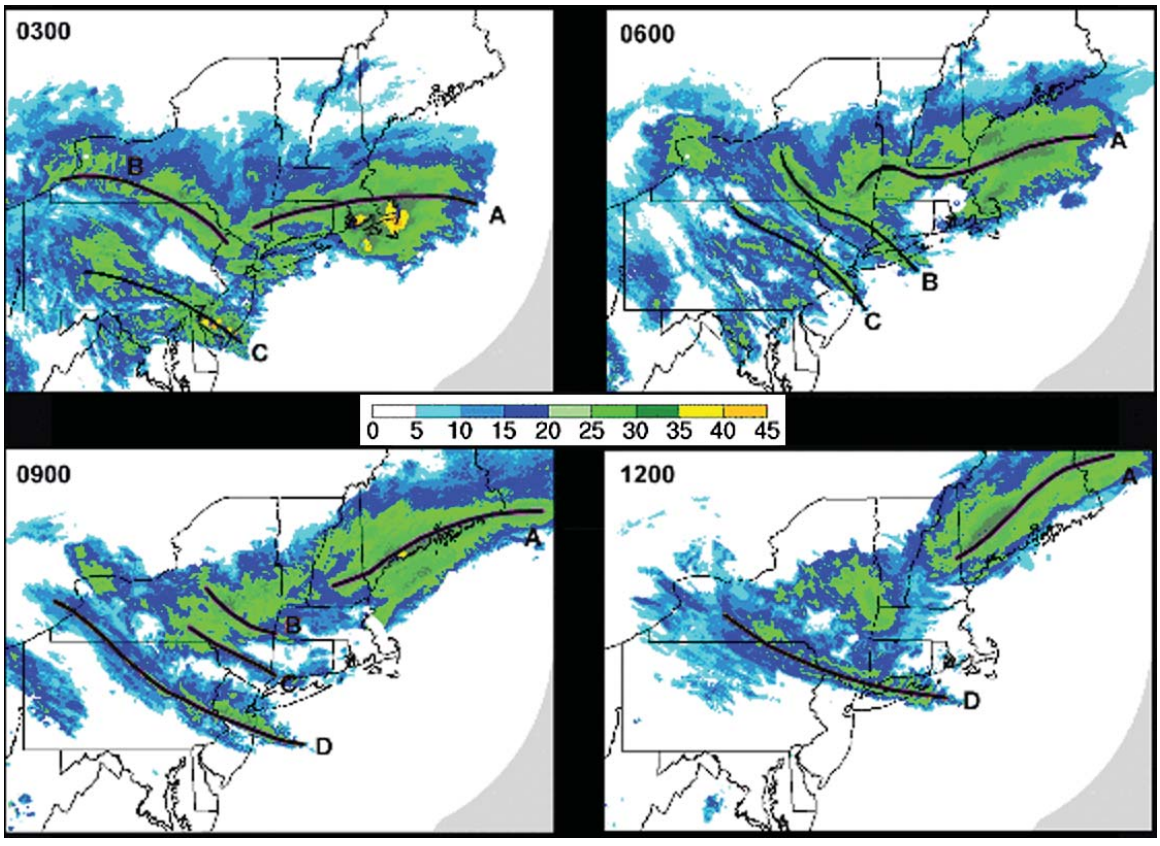


Fig. 5. Composite radar reflectivity (dBZ) at 0300, 0600, 0900, and 1200 UTC 1 Mar 2005, showing snowbands A, B, C, and D.

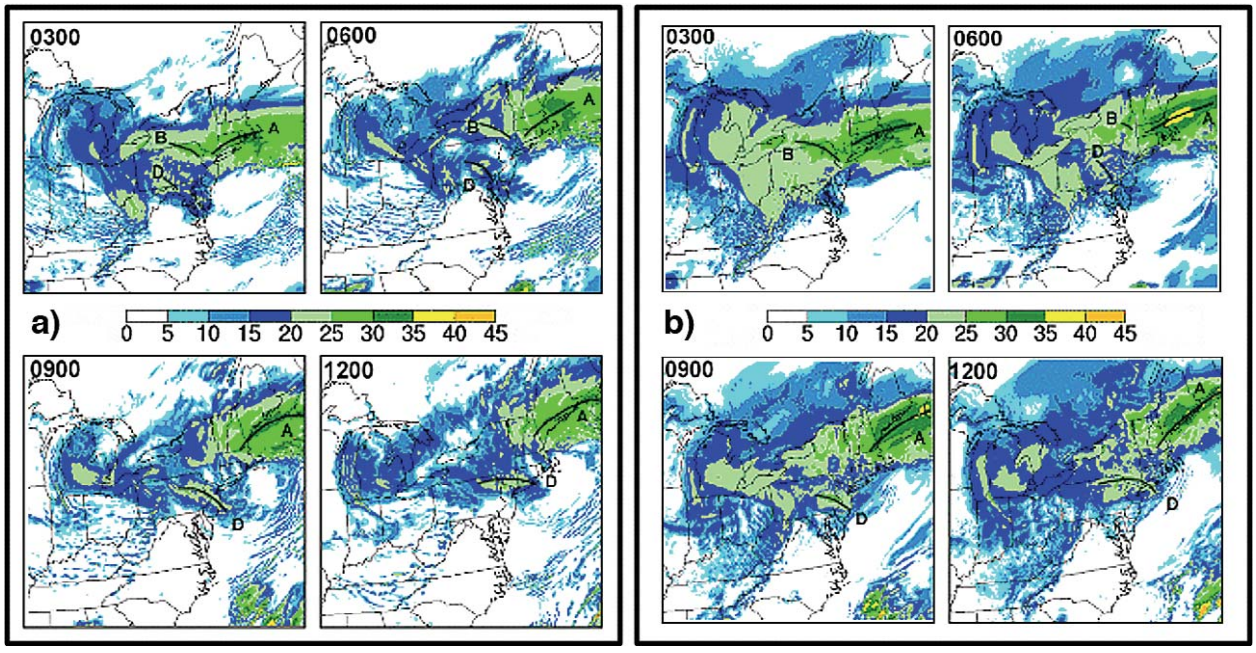


Fig. 6. Simulated composite radar reflectivity fields (dBZ) at 0300, 0600, 0900, and 1200 UTC 1 Mar forecast by the 0000 UTC 1 Mar 2005 cycle of the (a) ARW and (b) NMM, showing precipitation bands A, B, C, and D for comparison with Fig. 5.

is very limited in the 3-h accumulated precipitation fields (Fig. 4). The predictions of reflectivity can also be used in comparison between model forecasts and observed reflectivity to determine forecast performance in real time.

Storm-total precipitation predictions and verification (McDonald and Graziano 2001) (Fig. 7) show clearly that the WRF precipitation forecasts have more detail than either operational model. This trait is particularly apparent in New England, where both WRF models forecasted three distinct north–south bands of enhanced precipitation. The estimated precipitation does not display the same detail as shown in the forecasts, although there is a well-defined minimum north–south band across central Massachusetts in both the forecasts and the estimated precipitation. Snowfall reports in populated areas such as the Northeast can often have greater detail than precipitation observations. The snow accumulation, valid at the same time as the precipitation analyses shown in Fig. 8, depicts three north–south accumulation maxima, as seen in the DWFE forecasts (Fig. 7).

The Eta Model has some indication of a precipitation maximum in eastern New York, but misses the minimum in southern New England, while the GFS precipitation field is quite smooth and shifted too far east. In New England the amount of precipitation forecast by the WRF models is similar and more than either operational model. The estimated precipitation in the area does not verify the WRF model forecast amounts exceeding an inch, though based on the

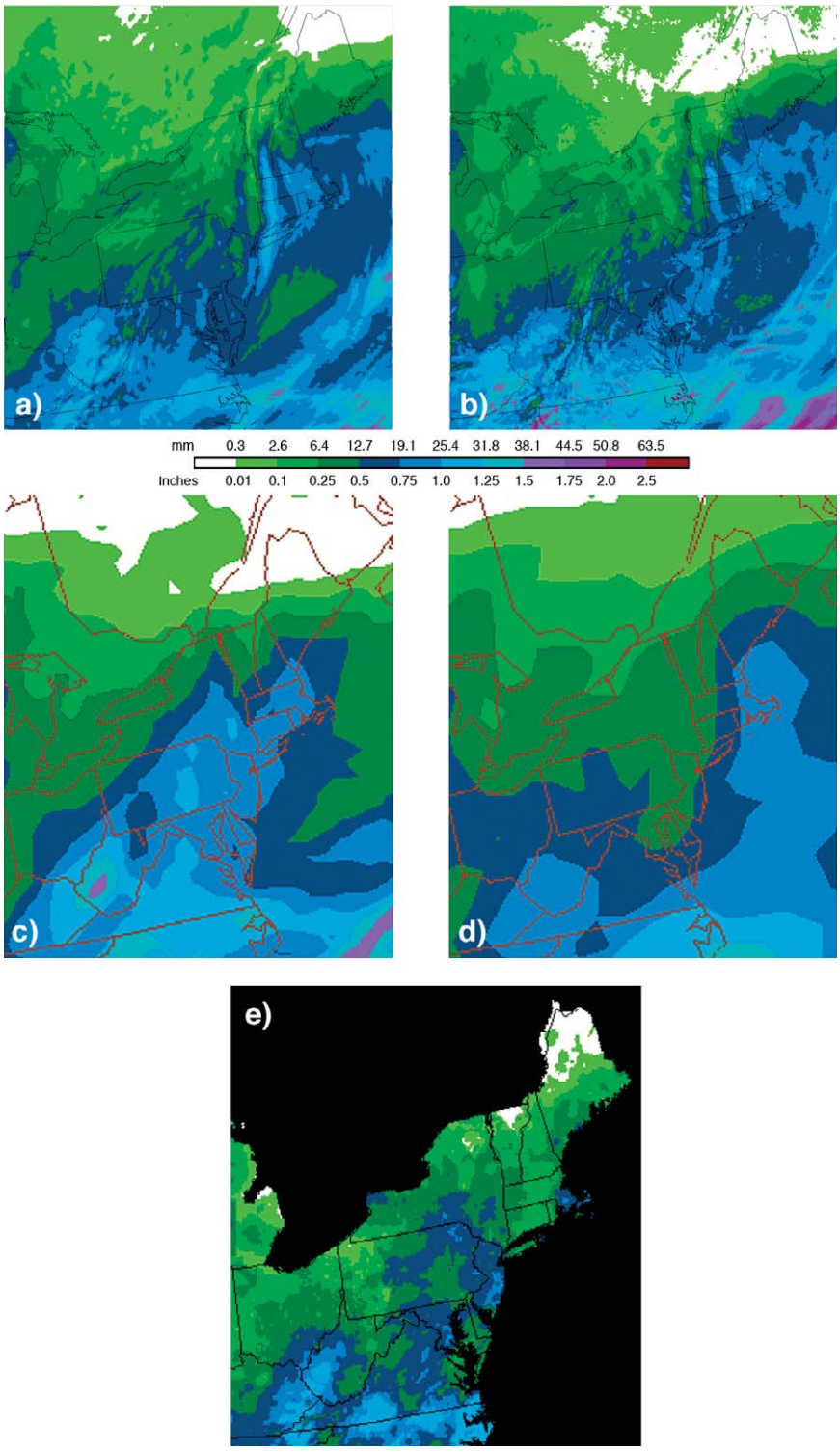


Fig. 7. Thirty-six-hour accumulated precipitation (in. and mm) at 1200 UTC 1 Mar 2005 from (a) ARW forecast, (b) NMM forecast, (c) NCEP operational Eta Model forecast, (d) NCEP operational GFS forecast, and (e) National Precipitation Verification Unit estimate.

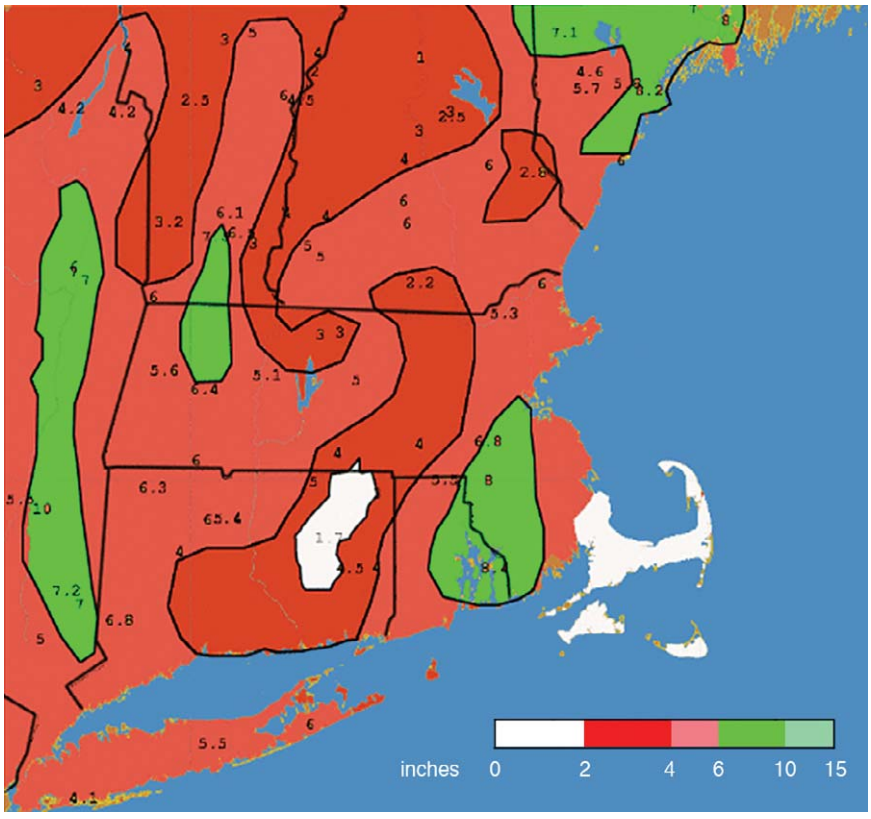


Fig. 8. Observed snow accumulation (in.) for the 24-h period ending at 1200 UTC 1 Mar 2005.

snowfall reports it is possible that local areas could have exceeded an inch of precipitation.

This case demonstrates the ability of the DWFE models to predict rather well the snowbands associated with one of a large number of snowstorms that affected the New England area during the winter of 2004/05. At the same time, this case illustrates how the two WRF forecasts are very close in appearance to each other, something seen repeatedly during DWFE. This characteristic suggests that the initial and boundary conditions were more influential in producing mesoscale structures than were the details of the numerics and physics in the two models.

Discussion. An important goal of the WRF DTC in setting up real-time forecast experiments is to obtain forecaster participation, which is not possible when runs are done retrospectively. The NWS Eastern, Southern, Central, and, to a lesser extent, Western Regions, participated in DWFE. The NWS participants were involved with NCAR, NOAA/ESRL, and NCEP in the planning, execution, and evaluation phases of DWFE. The requirements for real-time experiments differ from those for retrospective testing. Real-time experiments allow fewer failures because a few days of missing products can completely disengage forecasters, who have limited interest in looking for products that may not be readily available. Therefore, real-time experiments demand a higher reliability, which requires access to stable computer systems and robust workflow management scripts that attempt recovery when part of the end-to-end system fails (due to, e.g., input data being late or a computer or network going down). In addition, real-time experiments require that fields be accessible in a timely manner. One of the issues that limited the use of forecast products during DWFE was that neither NOAA/ESRL nor NCAR could run the models more than once a day. By comparison, the NAM and GFS are run 4 times a day, so a forecaster is more likely to use the latest guidance. In effect, the DWFE runs were "new" only for one of the three forecast shifts in a 24-h period. If possible, future real-time experiments should be run more frequently. Another limiting factor was that forecasters were extremely busy with the relatively new NWS IFPS, limiting their time devoted to DWFE products. Only a limited number of fields were available on AWIPS; additional fields had to be accessed through FX-Net, which was slow in offices with low bandwidth and did not allow the direct use of the forecast fields in IFPS. DWFE offered an opportunity to understand the challenges and assess the value of real-time experiments.

One area of much debate between developers, forecasters, and WRF DTC staff is the use of high-resolution, large-domain grids in the forecast process. Forecasters had difficulty using some of the primary fields supplied either by the WRF DTC or derived fields computed by AWIPS and FX-Net because they contained an excessive amount of detailed information. Forecasters do not have good tools to deal with such grids, especially when they cover the entire CONUS. High-resolution fields over the entire CONUS can be of little importance to a forecaster

whose focus is on the local scale of a NWS Forecast Office. Figure 9a shows a 12-h vertical velocity forecast by the ARW. Even though the forecast contains a lot of information, it is difficult to extract value from it because of the presence of a large number of mesoscale structures in the field. The WRF DTC attempted to address this problem by applying a nonlinear scale to the vertical velocity images displayed on the Web site, which enhanced the broader regions of ascending and descending motion, while retaining the local-scale details (Fig. 9b). This approach made the vertical velocity field more usable, but it is by no means a final answer, because it still produces a very busy plot.

The WRF DTC also received comments from forecasters regarding the vertical vorticity field. Again, this field contained a large number of small wavelength features, such as vorticity streamers (Koch et al. 2005), which made its interpretation challenging. Given that forecasters want to make synoptic-scale use of variables that have high variability in high-resolution forecasts (such as vertical motion and vorticity), DWFE has highlighted the need for the development of new tools to make these fields usable. Moreover, forecasters need to be educated and trained in the use of these new tools. The methods need to go beyond the use of smoothing; removing energy in the highest wavelengths while still retaining information on the scale of 100 km does not solve the problem of creating synoptic-scale fields. This is a problem of signal recovery that demands an investment of resources. Techniques involving the balance of large-scale and mesoscale flows might need to be employed to attempt a separation of scales. It is possible that a multiscale approach, similar to how NCEP transforms high-resolution NAM data into lower-resolution grids for use in large areas on AWIPS, might allow examination of both large-scale quasigeostrophic patterns as well as full resolution details associated with small-scale phenomena.

A novel field developed for DWFE was the simulated composite reflectivity. Although reflectivity has been produced for some other real-time forecast systems [such as the ARW forecasts generated by NCAR for the Storm Prediction Center National Severe

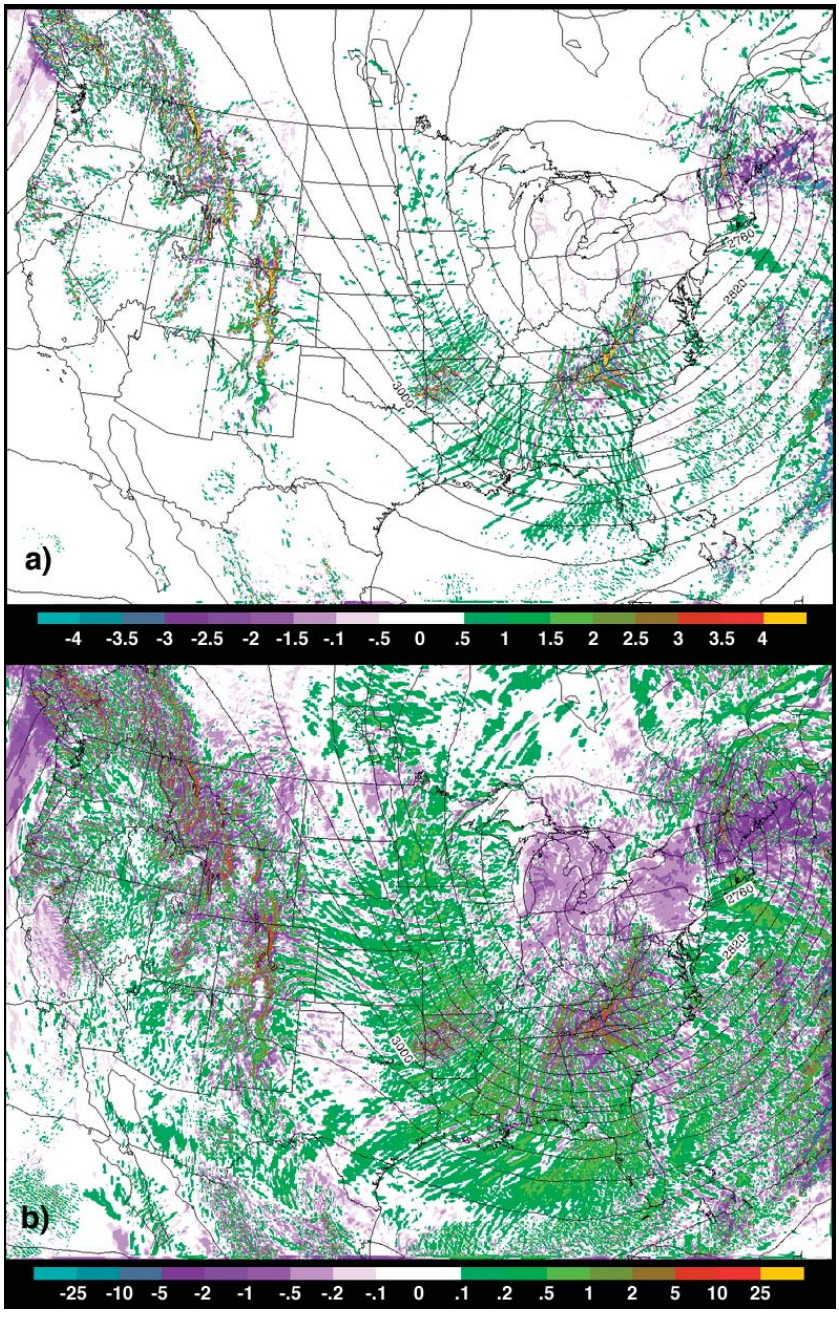


Fig. 9. Twelve-hour ARW forecast of 500-hPa vertical velocity (Pa s^{-1}) valid at 1200 UTC 1 Mar 2005 using (a) a linear scale and (b) a nonlinear scale.

Storm Laboratory Spring Program, as described by Kain et al. (2005)], this product was not routinely produced by NCEP at the time of DWFE, and was therefore not available for the NAM or GFS models. Forecasters found simulated composite reflectivity very useful because it could be employed to distinguish features not discernible in the 3-h accumulated precipitation forecasts, such as mesoscale snowbands and gravity waves (Koch et al. 2005). Additionally, simulated composite reflectivity can be directly compared with radar observations for verification (Fig. 5), even though an exact comparison is not possible because the radar observations are subject to distortions due to the Earth's curvature, bright bands, anomalous propagation, etc. Statistics showed that reflectivity was one of the most accessed fields on the DWFE Web site, only second to MSLP and 3-h accumulated precipitation. The popularity of the composite reflectivity display in DWFE motivated NCEP to include some reflectivity fields in certain NAM output files.

One area of much debate between DTC staff and forecasters was the computation of simulated reflectivity by the WPP. Reflectivity is a function of the number concentration of microphysical species, but the microphysical parameterizations used (Ferrier in the NMM and WSM-5 in the ARW) only predict the mixing ratio of the species. Therefore, a number of assumptions and a diagnostic method must be used to relate the number concentration to the mixing ratio. Such a relationship is necessary for the internal computation of the microphysics schemes (within WRF) and for the diagnostic computation of reflectivity (in the WPP). During DWFE, forecasters noticed subtle differences between the NMM- and ARW-simulated reflectivities. The NMM produced broader areas with low-reflectivity values (less than 25 dBZ) and did not reach the reflectivity peaks predicted by the ARW in convective cells (Fig. 6). Investigation of the WPP code revealed that the diagnostic relationship between the number concentration and mixing ratio used in the WPP was compatible with the Ferrier microphysics scheme [using the fixed-intercept method as described by Koch et al. (2006)], but did not match the WSM-5 scheme. Work done by the WRF DTC led to a recommendation to the WRF development group, subsequently adopted for at least one microphysics scheme, that simulated reflectivity needs to be computed within the WRF model microphysics and not by a separate program. This method assures reflectivity is computed in a manner consistent with the cloud microphysics scheme.

Because forecasters are very familiar with their own region of responsibility, their inspection of the high-resolution DWFE grids was valuable in determining problems with the forecasts. As an example, forecasters noticed that 2-m temperatures over the small lakes of Minnesota were abnormally high, contrasting with the much colder surroundings. The WRF DTC staff inspected these forecasts in more detail, because they had access to a larger number of variables than those supplied to forecasters. This inspection revealed that the lakes in the model were not frozen, which was not realistic in the winter season. The WRF DTC then contacted the model developers. They confirmed the problem and diagnosed that it originated with the method used in WRF to assign water temperature to lakes. WRF uses the nearest water point of the analysis that initializes the model. The analysis used was the Eta 212, which has a grid spacing of 40 km, and therefore cannot resolve most of the small lakes. In the Eta 212, the nearest water point to most of these small lakes was Lake Michigan, which was not frozen because of its large size and consequent high heat capacity. The identification of this weakness in WRF led to a development (which was later incorporated in WRF releases) to address the problem. Obviously, this problem could have been identified and reported by a researcher looking at forecasts in retrospective model runs. However, real-time experiments have the potential to increase the number of meteorologists looking at forecasts and, because forecasters are very experienced, the chances of systematic errors being detected are greater.

CONCLUDING REMARKS. In facilitating two-way transfer of information between the development and operational NWP communities, the WRF DTC has conducted both real-time and retrospective model evaluations. DWFE is an example of the importance of real-time experiments. While posing many technical challenges, real-time experiments incorporate valuable forecaster participation, increasing the number of people evaluating the forecasts and raising the chances that important model deficiencies are identified and corrected. Real-time experiments, however, do not preclude in-depth retrospective studies of the datasets produced. In fact, subsequent DWFE research has been performed by WRF DTC visitors (Skamarock and Dempsey 2005; Gallus and Bresch 2006).

One challenge of the real-time experiments is the handling of model errors and deficiencies detected during these experiments. Should such problems be corrected during the experiment, or should the

entire experiment be conducted with a frozen (but potentially inaccurate) model? From the perspective of forecasters, it is advantageous to correct errors during the experiment in order to receive the best possible guidance from the model. However, code modifications will compromise the forecasters' ability to evaluate the model if used too frequently. From the perspective of WRF DTC staff, which is trying to gather robust statistics to assess model behavior, a reasonably long period of runs with unaltered code is required. The WRF DTC's current recommendation is that extensive code testing be done in retrospective mode before beginning a real-time experiment, in order to eliminate most problems before the real-time period begins. Only major bugs should be corrected during a real-time experiment. If alterations in the code prevent the consolidation of statistical results, further data may be gathered in a complementary period of retrospective runs following a real-time experiment.

The statistics gathered during DWFE indicate that when two cores of the WRF model are configured similarly and initialized from similar initial conditions, the verification statistics produced are comparable for most variables. It is expected that, if the models were initialized in a cycled way, with each forecast's initial conditions depending on the previous cycle, the differences between the forecasts made with the two cores would be larger. As soon as the WRF DTC has a unified cycled initialization capability as part of its end-to-end system, the WRF DTC plans to run tests with that capability. With the current noncycled initialization, no significant differences between the ARW and NMM were found for the upper-air bias and RMSE of temperature and winds, and for the precipitation bias and ETS. The WRF model forecast verification scores are similar to those of the Eta Model for all variables except precipitation, indicating that the WRF model is producing forecasts of comparable skill to those of the NCEP operational suite. The precipitation bias scores differ vastly between the DWFE WRF models and the Eta Model, probably because the Eta Model runs at a lower resolution and makes use of the BMJ cumulus parameterization, which produces smoother precipitation forecasts.

DWFE has been useful in introducing the WRF model to NWS forecasters and in fostering exchanges between forecasters and model developers. One of the challenges of DWFE was the transfer of information to field offices in a timely manner. Forecast offices that used AWIPS to display DWFE output experienced better timing than those that employed FX-Net, a promising technology that was slow in offices with low bandwidth. Forecaster participation led to the correction of a bug in the model, and was useful in highlighting the need for further research on issues related to the display, distribution, and use of high-resolution models on the synoptic scale. As discussed by Koch et al. (2005) and Szoke et al. (2005), the forecasters indicated that DWFE products, especially the new simulated reflectivity field, added value to the forecasts of mesoscale winter weather phenomena, such as snowbands. In this sense, DWFE has been successful in bringing together the operational and research communities to join efforts in model evaluation, and in defining avenues for improving model performance and the use of models by forecasters.

The WRF DTC is continuing activities to bridge the gap between research and operational communities. Products from WRF DTC experiments and tests have been made available to the community for further investigations. Studies are being currently conducted at NCAR's Research Applications Laboratory using the WRF Rapid Refresh Core Test results. In 2007, the WRF DTC will be extending the WRF Core Test to a 60-h lead time. If the results indicate that the forecast skill is similar between the cores, research conducted with one core will be able to be directly transferred to another, therefore accelerating the rate of technology transfer to the NWS. In addition to transferring research results to operations, the WRF DTC will continue to support several codes run at NCEP for the general community. In addition to WRF NMM, already supported through the tutorials and the WRF help desk, the WRF DTC will soon begin support of the hurricane models (including ocean and wave models) and the GSI, used at NCEP as the analysis system for the NAM and Global Forecasting System, and selected for use with the upcoming WRF Rapid Refresh. Additionally, the WRF DTC has started (in 2007) the distribution and support of the MET, community software intended for traditional and advanced forecast verification. Last, WRF DTC will soon be making the operational configuration of the NAM available to the general community for testing and evaluation purposes.

ACKNOWLEDGMENTS. DWFE relied on invaluable participation from several institutions. From NCEP, Geoff DiMego and Nelson Seaman assisted with the planning, and Ying Ling supplied the NCEP QPFV System. Christopher Harrop, Jacques Middlecoff, Paul Hyder, and Robert Lipschutz, from NOAA/ESRL, and George Fuentes, Mark Gentry, and Juliana Rew from NCAR provided com-

putational support. From NOAA/ESRL, Linda Wharton and Paula McCaslin supported the use of AWIPS, while Sher Schranz and Jebb Stewart supported FX-NET. Kristin Conrad (NCAR) and Randall Collander (ESRL) supported the Web display. James Bresch from NCAR worked on software porting, computer infrastructure, and image display. Fei Chen, Kevin Manning, and Mukul Tewari from NCAR supplied the HRLDAS for the ARW. Barbara Brown from NCAR assisted with the significance testing. Dave Novak, Alex Tardy, Peter Browning, Jeff Waldstreicher, Josh Watson, Bernard Meisner, Mark Jackson, and Peter Manousos from the NWS assisted with the planning of the experiment and model evaluation. Pam Johnson provided administrative support. This article benefited from a technical review by Ann Reiser.

APPENDIX A: METHODOLOGY FOR FORECAST POSTPROCESSING AND VERIFICATION USED IN DWFE.

Postprocessing. The WPP software (Chuang et al. 2004; Chuang and Manikin 2001), developed at NCEP, was used. It serves three main purposes:

- to vertically interpolate the forecasts from their native vertical coordinate to isobaric levels,
- to compute derived meteorological fields, and
- to horizontally interpolate the forecasts from their native staggered grid to an unstaggered grid defined by the user.

During DWFE, the NMM and ARW forecasts were both horizontally interpolated to a common Lambert conformal grid (G163), with 5-km grid spacing covering an area very similar to the native grids of each model. Forecasts on G163 were used for dissemination to users and for verification.

The pressure reduction to sea level was computed in the WPP using the membrane method described in Chuang et al. (2004), while the reductions in temperature, RH, and wind to shelter level were computed by the WRF model itself and are compatible with the physics used in each setup. To meet the needs of the forecasters, new products, such as visibility, reflectivity, and precipitation type, were incorporated into the WPP. The last step in postprocessing was the conversion of the WPP from its native GRIB-1 format to GRIB-2. The significantly more compact nature of GRIB-2 was key for data transfers of large datasets.

Forecast verification. Forecasts were evaluated through subjective and objective verification. Subjective evaluation was conducted by the operational and research communities via online forms that are accessible through the DWFE catalog (Szoke et al. 2005), and informal communication with the NWS both during and following the experiment. The responses to the forms are summarized in Koch et al. (2005).

Objective verification employed two systems: the NCEP Verification System (Chuang et al. 2004) and RTVS (Mahoney et al. 2002; Loughe et al. 2001). Output of the verification from both systems was ingested into the RTVS database, which could be queried and graphically visualized through the RTVS Web-based interface. This approach facilitated the sharing of information among the experiment participants, who were distributed across a variety of locations. Easy access to verification statistics was valuable during the DWFE experiment, especially during early model development and testing.

Objective verification of precipitation. The DWFE QPFV made use of two approaches with the goal of comparing and contrasting the results obtained by the two methods (Demirtas et al. 2005a,b). For the results shown in this article, both approaches verified 24-h accumulations valid at 1200 UTC, corresponding to model forecast hours 12 through 36. The accumulation thresholds selected for verification were 0.01, 0.10, 0.25, 0.50, 0.75, 1.00, 1.50, and 2.00 in.

The NCEP QPFV system uses a grid-to-grid approach in which the forecasts and a gridded precipitation analysis are remapped onto a common grid and then compared. The grid used for comparison is the continental part of the 12-km grid spacing Grid 218 (G218), which encompasses the entire CONUS. This grid was selected because the NAM precipitation verification is routinely computed at NCEP for this grid. The remapping technique (Baldwin 2000; Accadia et al. 2003) maintains, to a desired accuracy, the total precipitation on the original grid. The NOAA Climate Prediction Center one-eighth-degree analysis, based on 7,000–8,000 daily gauge reports, was used as the observational dataset for verification using the NCEP Precipitation Verification System.

The RTVS QPFV developed at ESRL uses a grid-to-point approach to compare forecasts bilinearly interpolated to rain gauge locations against the quality-controlled Hydrometeorological Automated Data System. Approximately 4,500 stations are used daily. The verification covered an area that was virtually identical to the domain used in the NCEP QPFV, but the RTVS verification used the forecasts on the original G163 and not on G218.

The RTVS and NCEP QPFV systems compute a variety of scores. Only the area bias and the ETS

are discussed in this article. For further discussion regarding these scores, see Wilks (2005). The scores obtained for each forecast cycle are averaged in time to obtain mean statistics. The estimate of uncertainty in the QPFV was obtained by applying a numerical resampling method known as bootstrapping (Efron and Tibshirani 1994). This method utilizes different subsets of samples from the daily contingency tables with replacement. By accumulating the counts of the daily contingency tables over the entire DWFE period, a single new contingency table for each core/threshold is derived. The bias and ETS are then calculated using this new contingency table. To obtain 95% confidence intervals, this process is repeated a large number of times (5,000 in this case), yielding an empirical distribution of the statistics, from which the uncertainty in the statistic can be estimated by finding the confidence bounds that separate 2.5% of the 5,000 empirical samples into each tail.

Objective verification of temperature, relative humidity, and winds. Objective verification of upper-air and surface fields, except precipitation, was performed using just the NCEP verification system. This package employs a grid-to-point verification approach in which forecast fields are bilinearly interpolated to station location.

At upper levels, forecasts of temperature, RH, and winds were compared against rawinsondes to generate bias and RMSE averaged in time over the entire DWFE period. At the surface, forecasts of MSLP, 2-m temperature, 2-m RH, and 10-m winds were compared against METAR observations. The resulting statistics are averages of independent measures, because the autocorrelations between forecast statistics were examined and not found to be significant. Thus, 95% confidence intervals could be calculated for each statistic based on a standard algorithm for normal distributions that takes into consideration the variance of the sample (Wilks 2005). These confidence intervals give an estimate of the uncertainty in the statistics due to sampling error. Other sources of uncertainty, such as error in the observations or seasonal variability, were not considered. Further, these confidence intervals give an indication of the statistical significance of results, which is necessary but not sufficient to indicate practical significance. This is especially true when sample sizes are large, because most results tend to be statistically significant but may not have weather forecasting significance.

APPENDIX B: LIST OF ACRONYMS.

ARW	Advanced Research Weather Research and Forecasting model
AWIPS	Advanced Weather Information Processing System
BMJ	Betts–Miller–Janjic
CI	Confidence interval
CONUS	Continental United States
DTC	Developmental Testbed Center
DWFE	DTC Winter Forecasting Experiment
EOL	Earth Observing Laboratory
ESRL	Earth System Research Laboratory
ETS	Equitable threat score
GFDL	Geophysical Fluid Dynamics Laboratory
GSI	Gridpoint Statistical Indicator
HRLDAS	High-Resolution Land Data Assimilation System
HRW	High-resolution window
IFPS	Interactive Forecast Preparation System
LSM	Land surface model
MET	Model Evaluation Tool
MSLP	Mean sea level pressure
NAM	North American Mesoscale Model
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NMM	Nonhydrostatic mesoscale model
NMM5-CONUS	Nonhydrostatic mesoscale model 5-km grid spacing configured over the continental United States
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical weather prediction

NWS	National Weather Service
QPFV	Quantitative Precipitation Forecast Verification
RH	Relative humidity
RMSE	Root-mean-square error
RRTM	Rapid radiative transfer model
RTVS	Real Time Verification System
USWRP	U.S. Weather Research Program
UTC	Coordinated Universal Time
WFO	Weather Forecast Office
WPP	WRF Model Postprocessor
WRF	Weather Research and Forecasting
WSM-5	WRF Single Moment 5-class microphysics parameterization

REFERENCES

Accadia, C., S. Mariani, M. Casaioli, A. Lavagnini, and A. Speranza, 2003: Sensitivity of precipitation forecast skill scores to bi-linear interpolation and a simple nearest-neighbor average method on high-resolution verification grids. *Wea. Forecasting,* **18,** 918–932.

Baldwin, M. E., cited 2000: QPF verification documentation. NCEP/EMC. [Available online at www.emc.ncep.noaa.gov/mmb/ylin/pcpverif/scores/docs/mb-doc/pptmethod.html.]

Bernardet, L., L. Nance, H. Chuang, A. Loughe, and S. Koch, 2004: Verification statistics for the NCEP WRF pre-implementation test. Part 1: Deterministic verification of ensemble members. Preprints, *2004 WRF/MM5 Joint Workshop,* Boulder, CO, NCAR. [Available online at www.mmm.ucar.edu/mm5/workshop/ws04/PosterSession/Bernardet.Ligia_part1.pdf.]

——, P. Bogenschutz, J. Snook, and A. Loughe, 2005: WRF forecasts over the southeast United States: Does a larger domain lead to better results? Preprints, *Sixth WRF User's Workshop,* Boulder, CO, NCAR. [Available online at www.mmm.ucar.edu/wrf/users/workshops/WS2005/abstracts/Session2/10-Bernardet.pdf.]

Black, T., 1994: The new NMC mesoscale Eta model: Description and forecast examples. *Wea. Forecasting,* **9,** 265–278.

Bosart, L. F., and F. Sanders, 1986: Mesoscale structure in the megalopolitan snowstorm of 11–12 February 1983. Part III: A large amplitude gravity wave. *J. Atmos. Sci.,* **43,** 924–939.

Brown, J. M., S. Benjamin, T. G. Smirnova, G. A. Grell, L. R. Bernardet, L. B. Nance, R. S. Collander, and C. W. Harrop, 2007: Rapid-Refresh core test: Aspects of WRF-NMM and WRF-ARW forecast performance relevant to the Rapid-Refresh application. Preprints, *22nd Conf. on Weather Analysis and Forecasting,* Park City, UT, Amer. Meteor. Soc., 4A.1. [Available online at http://ams.confex.com/ams/pdfpapers/124822.pdf.]

Chen, F., K. W. Manning, D. N. Yates, M. A. Lemone, S. B. Trier, R. Cuena, and D. Niyogi, 2004: Development of High-Resolution Land Data Assimilation System and its application to WRF. Preprints, *16th Conf. on Numerical Weather Prediction,* Seattle, WA, Amer. Meteor. Soc., 22.3. [Available online at http://ams.confex.com/ams/pdfpapers/67333.pdf.]

Chuang, H., and G. Manikin, 2001: The NCEP Meso Eta model post processor: A documentation. NOAA/NWS/NCEP Office Note 438, 52 pp. [Available online at wwwt.emc.ncep.noaa.gov/mmb/papers/chuang/1/OF438.html.]

——, G. DiMego, M. Baldwin, and WRF WRF DTC Team, 2004: NCEP's WRF post-processor and verification systems. Preprints, *Fifth WRF/14th MM5 Users' Workshop,* Boulder, CO, NCAR. [Available online at www.mmm.ucar.edu/mm5/workshop/ws04/Session7/Chuang.Hui-Ya.pdf.]

Davis, C., L. Nance, L. Bernardet, M. Pyle, and H. Chuang, 2004: WRF forecasts of recent major weather events: A comparison of EM and NMM cores. Preprints, *2004 WRF User's Workshop,* Boulder, CO, NCAR. [Available online at www.mmm.ucar.edu/mm5/workshop/ws04/PosterSession/Davis_Chris2.pdf.]

Demirtas, M., L. Nance, L. Bernardet, Y. Lin, H. Chuang, A. Loughe, R. Gall, and S. Koch, 2005a: Verification systems used for the Developmental Tested Center. *Sixth WRF/15th MM5 Users' Workshop,* Boulder, CO, NCAR, P3.23.

——, and Coauthors, 2005b: Quantitative precipitation forecast (QPF) verification of DWFE. Preprints, *21st Conf. on Weather Analysis and Forecasting,* Washington, D.C., Amer. Meteor. Soc.,

7.2. [Available online at http://ams.confex.com/ams/pdfpapers/94945.pdf.]

Done, J., C. Davis, and M. Weissman, 2004: The next generation of NWP: Explicit forecasts of convection using the Weather Research and Foreasting (WRF) model. *Atmos. Sci. Lett.,* **5,** 110–117.

DTC, cited 2006: DTC visitor reports for FY 05/06. [Available from Developmental Testbed Center, NCAR, P.O. Box 3000, Boulder, CO, 80307-3000, or online at www.dtcenter.org/visitor_program/visitors.05_06/agenda_presentations.php.]

Efron, B., and R. Tibshirani, 1994: *An Introduction to the Bootstrap.* Chapman and Hall/CRC, 436 pp.

Gallus, W., and J. Bresch, 2006: Comparison of impacts of WRF dynamic core, physics package, and initial conditions on warm season rainfall forecasts. *Mon. Wea. Rev.,* **134,** 2632–2641.

Janjic, Z. I., 2003: A nonhydrostatic model based on a new approach. *Meteor. Atmos. Phys.,* **82,** 271–285.

——, 2004: The NCEP WRF core. Preprints, *20th Conf. on Numerical Weather Prediction,* Seattle, WA, Amer. Meteor. Soc., 12.7. [Available online at http://ams.confex.com/ams/pdfpapers/70036.pdf.]

Jankov, I., W. A. Gallus Jr., M. Segal, B. Shaw, and S. E. Koch, 2005: The impacts of different physical parameterizations and their interactions on warm season MCS rainfall. *Wea. Forecasting,* **20,** 1048–1060.

Kain, J., S. Weiss, M. Baldwin, G. Carbin, D. Bright, J. Levit, and J. Hart, 2005: Evaluating high-resolution configurations of the WRF model that are used to forecast severe convective weather: The 2005 SPC/NSSL spring experiment. Preprints, *21st Conf. on Weather Analysis and Forecasting,* Washington, D.C., Amer. Meteor. Soc., 2A.5. [Available online at http://ams.confex.com/ams/pdfpapers/94843.pdf.]

Koch, S., and Coauthors, 2005: Lessons learned from the WRF DTC winter forecast experiment. Preprints, *21st Conf. on Weather Analysis and Forecasting,* Washington, D.C., Amer. Meteor. Soc., 7.6. [Available online at http://ams.confex.com/ams/pdfpapers/94854.pdf.]

——, M. Stoelinga, B. Ferrier, M. Demirtas, and J. Kain, 2006: Verification of convection-resolving model forecasts: Accumulated precipitation vs. simulated radar reflectivity fields. *Second Int. Symp. Quantitative Precipitation Forecasting and Hydrology,* Boulder, CO, WMO.

Loughe, A., J. Henderson, J. Mahoney, and E. Tollerud, 2001: A verifiction approach suitable for assessing the quality of model-based precipitation forecasts during extreme precipitation events. Preprints, *Symp. on Precipitation Extremes: Prediction, Impacts and Responses,* Albuquerque, NM, Amer. Meteor. Soc., P1.36.

Mahoney, J. L., J. K. Henderson, B. G. Brown, J. E. Hart, A. Loughe, C. Fisher, and B. Sigren, 2002: The Real-Time Verification System (RTVS) and its application to aviation weather forecasting. Preprints, *10th Conf. on Aviation, Range and Aerospace Meteorology,* Portland, OR, Amer. Meteor. Soc., 9.8. [Available online at http://ams.confex.com/ams/pdfpapers/40728.pdf.]

McDonald, B. E., and T. G. Graziano, 2001: The national precipitation verification unit (NPVU): Operational implementation. Preprints, *Symp. on Precipitation Prediction: Extreme Events and Mitigation,* Albuquerque, NM, Amer. Meteor. Soc., P1.32.

Nance, L., 2006: Weather Research and Forecasting core test: WRF DTC report. NCAR, 36 pp. [Available from Developmental Testbed Center, NCAR, P.O. Box 3000, Boulder, CO, 80307-3000, or online at http://ruc.fsl.noaa.gov/coretest2/WRF DTC_report.pdf.]

Seaman, N., and Coauthors, 2004: The WRF process: Streamlining the transition of new science from research into operations. Preprints, *Fifth WRF User's Workshop,* Boulder, CO, NCAR.

Skamarock, W. C., and D. Dempsey, 2005: The WRF DTC winter forecast experiment: Analysis of ARW and NMM. Preprints, *Seventh WRF Users' Workshop,* Boulder, CO, NCAR.

——, J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2005: A description of the advanced research WRF version 2. NCAR Tech Note NCAR/TN-468+STR, 88 pp.

Szoke, E., S. Koch, and D. Novak, 2005: An examination of the performance of two high-resolution numerical models for forecasting extended snow bands during the DTC Winter Forecasting Experiment. Preprints, *21st Conf. Weather Analysis and Forecasting,* Washington, D.C., Amer. Meteor. Soc., J1.36. [Available online at http://ams.confex.com/ams/pdfpapers/94473.pdf.]

Wang, N., S. Madine, and R. Brummer, 2002: Investigation of data compression techniques applied to AWIPS datasets. Preprints, *Interactive Symp. on the Advanced Weather Interactive Processing System (AWIPS),* Orlando, FL, Amer. Meteor. Soc., J261–J263. [Available online at http://ams.confex.com/ams/pdfpapers/27337.pdf.]

Wilks, D. S., 2005: *Statistical Methods in the Atmospheric Sciences.* 2nd ed. Academic Press, 648 pp.

Zhang, F., S. E. Koch, C. A. Davis, and M. L. Kaplan, 2001: Wavelet analysis and the governing dynamics of a large-amplitude mesoscale gravity wave event along the East Coast of the United States. *Quart. J. Roy. Meteor. Soc.,* **127,** 2209–2245.